

# The Moral Risks of Risky Technologies

Peter Goldie

## 1 Introduction

In his *Civilization and its Discontents*, Freud, writing in 1930, noted our increasing dependence on technologies – ships, aircraft, spectacles, telescopes, cameras, gramophones, telephones, and so on. He said, “Man has, as it were, become a kind of prosthetic God. When he puts on all his auxiliary organs he is truly magnificent; but those organs have not grown on to him and they still give him much trouble at times. Nevertheless, he is entitled to console himself with the thought that development will not come to an end precisely with the year 1930 A.D. Future ages will bring with them new and probably unimaginably great advances in this field of civilization and will increase man’s likeness to God still more. But in the interests of our investigations, we will not forget that present-day man does not feel happy in his Godlike character” (Freud 1930/1985, pp. 279–280).

Since 1930, there have indeed been “unimaginably great advances” in technologies: computers, satellites, GPS navigation systems, mobile telephones, robots, embodied conversational agents (ECAs), avatars, androids, and so on. But our auxiliary organs still “give us much trouble at times”: they go wrong; they take on a life of their own; and they are often incomprehensible in their function.

So in spite of these advances we still do not feel entirely happy in our “Godlike character”. Quite what the unhappiness consists of is not entirely clear. This is an empirical issue on which I do not wish to reach any definitive conclusions. My main interest here will be in the normative issues. However, it would seem that, to a considerable extent, we have ambivalent or mixed feelings towards our auxiliary organs – and these are manifested in particular in our emotional responses towards them, and in the other ways in which we interact with them. From now on I will limit my discussion of technologies to computers, robots, avatars, ECAs, and other

---

P. Goldie (✉)

Department of Philosophy, University of Manchester, Manchester, UK  
e-mail: peter.goldie@manchester.ac.uk

kinds of emotion-oriented technologies (EOTs), many of which are known as semi-intelligent information filters (SIIFs); in general, these are kinds of technologies with which we tend to *interact* (often emotionally), and not just act towards, as we do, for example, towards those which Freud discusses. It is largely for this reason that the interesting issues which I want to discuss here arise.

## 2 Ambivalence in Our Behaviour Towards Technologies

On the one hand, it seems that we relate to these technologies as if they are simply what they in fact are: inanimate objects, incapable of any kind of thought or feeling, and thus no more deserving of any kind of *human* interaction as might be a screw-driver or a cabbage. And yet even here, when they “give us trouble”, we verbally and physically abuse them in ways that are somehow oddly expressive of our frustration (de Angeli et al. 2006). For example, we complain that the thing has a “mind of its own”, we shout and swear at it (“Come on, you damned thing, work!”), and we beat it, often to our own detriment as well as to the machine itself. These kinds of actions are clearly expressive of emotions such as frustration and anger (Hursthouse 1991), but the manner of expression is in many ways peculiar to this kind of object: we are, for example, less likely to shout at a tube of toothpaste if it fails to work, whereas it is typical behaviour towards a malfunctioning computer, or to the pre-recorded telephone message from the airline company, telling us they are “sorry” to keep us waiting, and that our call is “valuable” to them.

Ambivalence is revealed in other empirical findings that we are also (and somewhat contradictorily to our aggressive behaviour) capable of behaving politely towards computers, unconsciously treating them as we might humans, whilst at the same time denying that we think of them as human (Nass and Reeves 1996).

Yet further evidence of ambivalence is found in studies which have involved carrying out Milgram-style experiments on what are known by participants to be inanimate avatars and robots (Milgram 1974; Slater et al. 2006; Rosalia et al. 2005; Bartneck et al. 2006; Bartneck and Hu 2008). In one series of experiments on an avatar, a female virtual person, the investigators concluded as follows: “Our results show that in spite of the fact that all participants knew for sure that neither the stranger [the avatar] nor the shocks were real, the participants who saw and heard her tended to respond to the situation at the subjective, behavioural and physiological levels as if it were real” (Slater et al. 2006, p. 1).

And finally, of course, there is the vexed question of what to make of the “uncanny valley”, as introduced by Masahiro Mori (Mori 1970), and now much discussed in robotics and computer science. What Mori argued was that our emotional attitudes towards robots change as the robots become more and more similar to human beings (in behaviour, in facial and verbal expression, and so on). We are thus more comfortable with a humanoid robot than an industrial robot, and yet when the robot becomes even closer in appearance to a healthy human but is still clearly *not* human, our feelings of comfort and familiarity decline: we are in the uncanny valley. There are a number of explanations that have been put forward for this kind

of reaction that we have: that the robots are “bukimi” in Mori’s sense – weird, ominous, eery; that they give rise to disgust; that they deviate from the norms of physical beauty; that they frustrate our (largely unconscious) expectations; that they give rise to fear of death (MacDorman and Ishiguro 2006).

These emotional responses and patterns of behaviour, expressive of our ambivalence, are generally not of the kind that can be seen as rational, in the way that, for example, fear of a savage dog would be rational. They are, rather, more visceral, more primitive.

### 3 Not Ambivalence in Belief

It is important to appreciate here that this ambivalence in our emotional responses and behaviour does not seem in any way to be grounded in ambivalence in our *beliefs* about whether or not computers, robots, and so on are minded, and thus capable of thoughts and feelings. The point can be put in terms of the more general contrast between two kinds of consciousness: what the philosopher Ned Block (1997) has called *access consciousness*, as contrasted with *phenomenal consciousness*. Roughly, access consciousness is the kind of consciousness involved in mere cognition – information storage and processing for example. So, for example, the capacity of something to recognise a threat and to respond with evasive behaviour has access consciousness. And, still as part of access consciousness, a more complex organism might also be capable of recognising its own internal states, such as the state which represents *that* it is threatened and *that* a certain kind of evasive response is called for. Phenomenal consciousness, in contrast, is what is involved when *there is something that it is like* for the organism – in this case, where there is something that it is like to feel fear (Nagel 1974). There is something that it is like to be a human, a dog, or a cow – they all have phenomenal consciousness, and they all can experience fear – but there is nothing it is like to be a computer, or a robot.

Could there ever be something that it is like to be a robot – could a robot ever, for example, experience fear? As science fiction literature and film attest, we feel unsettled by the apparent fact that, in the fiction, these non-animal things are capable of emotional feelings, and we feel inclined to empathise with them in this respect. Consider, for example, the Nexus-6 replicants in *Blade Runner* (Ridley Scott 1982) who are programmed with a fail-safe device to cease functioning after 4 years in case they start to develop empathy; and the computer Hal in *2001: A Space Odyssey* (Stanley Kubrick 1968), which seems to be motivated emotionally, by revenge or envy perhaps, and seems to suffer as his systems are shut down. But these are thought experiments, and there is no evidence that adults are inclined to believe that actual technologies are capable of experiencing emotions (Picard 2002). The ambivalence in our behaviour and emotional responses, and even in our empathetic responses on some occasions, does not then seem to be grounded in an ambivalence or uncertainty in belief. And this stands in marked contrast to how we might, for

example, be ambivalent or uncertain in our beliefs about what is going on in the struggling trout on the end of the fishing line, or in the harpooned whale in its final death throes; in such cases, we might indeed be unsure of what is going on in the living creature.

So far, then, the discussion has been restricted to the empirical question of what our attitudes and behaviour are towards technologies of the kinds I have been focusing on, and my supposition is that these involve ambivalence of emotion and behaviour, but not ambivalence or uncertainty of belief.

Be that as it may, it is to the normative question that I now want to turn, and this will be the focus for the remainder of this chapter. What sort of attitudes and behaviour *ought* we to adopt towards these technologies?

## 4 The Rationality of Our Responses to Technologies

One obvious thought might be suggested to begin with: whatever else our attitudes and behaviour ought to be, they ought at least to be rational. However, on examination this thought runs the risk of proving either too much or too little. The point can be made by reference to a parallel argument in relation to our emotional engagement with fictional characters, an argument which is supposed to reveal a paradox of irrationality – the so-called paradox of fiction. It is paradoxical that each of these three propositions is intuitively acceptable: that we feel emotions towards fictional characters; that to be rational in feeling an emotion towards something we must believe that thing to exist; and yet we do not believe that fictional characters exist. Colin Radford, for example, has argued extensively that there is no acceptable reply to this paradox, and that it shows that our emotional responses to fictional characters are irrational: inconsistent and so incoherent (Radford 2001).

This conclusion, if true, would surely prove too much if it showed that we ought not to have emotional responses to fictional characters, simply on the grounds that such responses are irrational. And it would prove too little if it showed only that we have manifested a form of irrationality, without any implication that it ought not to be encouraged in other respects. In my view – which I cannot argue for here – the central difficulty with the so-called paradox of fiction is that the notion of rationality that is at work in setting up the paradox is so thin (Goldie 2009) that it has little force in recommending how we ought, all things considered, to think and feel.

It can be readily seen how a similar paradox could be set up for our emotional responses to, for example, the “cruel” treatment of an avatar of the kind found in the Milgram-style experiments that I mentioned above. The paradox would go something like this: we respond (let us assume) with moral concern to the treatment of the robot; we ought rationally to respond with moral concern to the treatment of something only if we believe that thing to have thoughts and feelings; and yet we do not believe that robots have thoughts and feelings. This argument might indeed show that this kind of response is irrational, but still, as with the parallel

argument about our emotional engagement with fictional characters, it either shows too much or too little. What we need to do is to consider the wider normative considerations, both moral and practical, that enter into addressing the question of how we ought, all things considered, to relate to technologies of the kinds I am concerned with.

## 5 Instrumental and Non-instrumental Value

Some things have merely instrumental value: something which is of instrumental value is to be valued only in so far as it is good of its kind, so that it performs its function well. For example, a knife is instrumentally valuable only in so far as it is able to cut; once it ceases to be able to perform that function, it ceases to be of value.

Shocking as it might be to us, Aristotle thought that slaves were valuable only in this way: “The slave”, he said, “is a living tool” (*Nicomachean Ethics*, 1161 b 4). But we should not, in recoil from this, turn to rejecting the idea that people should ever be thought of as having instrumental value. For it is undeniable that the taxi-driver, the housekeeper, the nanny, the man in the ticket office, can all have this kind of value. Rather, we should accept that humans can have instrumental value, but we should at the same time insist that they also have non-instrumental value, that they are of value for themselves, and not only for some further purpose. This is what is behind the “merely” in Kant’s famous claim: “So act that you always treat humanity . . . always at the same time as an end, never merely as a means” (1785/1964, p. 429).

It is controversial quite what is involved in treating people as ends, but I do not need to appeal to anything more here than a negative duty which is at least part of what is involved: the duty not to abuse people, not to treat them cruelly or aggressively. Of course more than that is involved in how we ought to treat people, but this will not be my concern here, for reasons which will emerge.

There is no doubt that technologies have instrumental value – when they work. The question that is pressing is whether, like people, they also have non-instrumental value, and if so, of what kind. There is, in fact, a range of possible sources of non-instrumental value here: we do not have to attribute non-instrumental value to technologies for the same reasons – essentially moral reasons – as we have to attribute this kind of value to people. I will briefly consider three other possible sources of non-instrumental value before turning to moral reasons of the kind that Kant had in mind.

One possible source of non-instrumental value that might apply to something such as a tool or a piece of technology is sentimental value (Hatzimoyisis 2003). For example, if I have a fountain pen that was given to me by someone I hold very dear, then I might well continue to treasure that pen even after it has ceased to perform its function well – even after it no longer works. There is no doubt that technological things sometimes do have sentimental value in this way: for example, some people

hang on to old and highly unreliable laptops just because they now have this kind of value for them. But it should be noticed about this kind of value that the value depends on the existence of the relevant associations, and it follows that the value is agent-relative in the sense that something which is of sentimental value for me need not be of sentimental value for you, just because it does not possess the relevant associations for you.

A second possible source of non-instrumental value of technologies is that one comes to consider them to be, in some sense, friends or companions. (For discussion of the value of friendship, see Stocker (1976).) Again, there are no doubt instances of this to be found, such as the way children behave towards their Tagamochi toys. But this value, like sentimental value, is agent-relative, and, moreover, there are perhaps concerns to do with the possibility of psychic disharmony that might undermine this kind of attitude. (Note here that I make the point not in terms of irrationality, but in wider terms to do with possible damage to the individual.)

Thirdly, there is aesthetic value, which, unlike sentimental value and value as friends or companions, is not agent-relative. A distinction of Kant's here is helpful in distinguishing two ways in which a piece of technology might have aesthetic value. Kant, in his great work on aesthetics, *The Critique of Judgement* (1790/1953), distinguished between free and dependent beauty. As Kant put it, "The first presupposes no concept of what the object should be; the second does presuppose such a concept and, with it, an answering perfection of the object" (Kant 1790/1953, p. 72). Interpretation is famously tricky here, but the essential idea is that something is freely beautiful if we can judge it to be beautiful without having a clear idea of what kind of thing it is or what its purpose is; Kant's example was the beauty of a flower. In contrast, something is dependently beautiful if we need to know what kind of thing it is, and what its purpose is, before we can judge its beauty; as Kant says, it is "ascribed to Objects which come under the concept of a particular end" (Kant 1790/1953, p. 72). For example, we might need to know that something is a rapier, and what the purpose of a rapier is, in order to judge its beauty: our judgement depends on this prior knowledge. Kant's own examples included men, horses, and buildings (Scarre 1981).

It strikes me that pieces of technology are capable of possessing either or both of these kinds of aesthetic value. The enormous NASA computer facility containing cabinet after cabinet of quietly humming mainframes might possess dependent beauty, because we need to know that the purpose of this facility is to track the movement of the stars in the Solar System if we are to appreciate its beauty. In contrast, perhaps the latest Apple laptop is freely beautiful: its design is such that we can admire its beauty without first needing to know what it is or what is its purpose.

So there are these three kinds of reasons for attributing non-instrumental value to technologies. Each of them is, I think, interesting in its own right, and may well have application in particular cases, but what I am seeking is a kind of reason that is somewhat more universal in its application than these, and with that in mind I now turn to moral considerations.

## 6 Moral Reasons for Valuing Technologies

Why might we think that technologies have moral value of a kind which is non-instrumental, so that they are valuable not only for some further purpose? Again, there are a number of possibilities here, and I want to eliminate some before turning to what I think is the most important moral consideration.

First, we might think that technological items such as robots have rights. Peter Singer has argued for a number of years that non-human animals have rights (for example, in Singer 1977), and it has even been suggested recently (in a report titled “Robo-Rights” commissioned by the UK Office of Science and Innovation’s Horizon Scanning Centre in December 2006) that rights could indeed be extended to robots. But even if we reject that idea as sheer madness (and the report was highly criticized at the time), we might still think that we have duties towards them. More interesting, though, is the thought that we have duties *with regard to* them, and it is this thought that I will turn to later.

Secondly, we might think that we should attribute moral value to robots and so on because they are sentient, or at least because we are not certain whether or not they are sentient, and we should, so to speak, give them the benefit of the doubt. But this is something that I considered earlier. We do not believe that they are sentient, and we do not seem even to believe there to be any doubt about the matter, so no moral choice arises here, as it might with fish or whales for example (Dennett 1996), even if we do sometimes empathise with them as if they are sentient. And it seems to me that we are right about this. Leaving aside any science-fiction future possibilities, we are in fact right to believe the contrary: to believe that current technologies do not possess phenomenal consciousness.

Even so, perhaps we should attribute moral value to them at least on the grounds that they do seem to possess *intelligence*, in the sense that they seem to possess *access* consciousness (Bartneck et al. 2006). Intelligence could be something that we should value in the world not only for its instrumental value. Perhaps, but I will leave that interesting thought, like the others, in suspense in order to turn to the moral considerations that I think bear most weight here, and has the widest range of application to technologies beyond just robots, computers, and other technologies that seem to possess intelligence.

Here is the central idea. The way we treat technologies can be expressive of our personality. Consider, for example, the person who regularly shouts at his computer for not working as he wants, bashing the “Enter” key in irritation and frustration. What might begin as behaviour towards just this computer can easily become more general and expressive of personality traits, such as irritability and short-temperdness, directed towards a wide range of technologies: towards the computer, towards the ticket machine in the railway station, towards the airline’s automatic telephone answering system, and so on. This irritable and short-tempered behaviour can then easily become generalised beyond technologies to people as well: towards the person in the ticket office as well as towards the ticket machine; towards the airline official on the telephone as well as towards the automatic answering system.

These officials come no longer to be treated with the respect that should be accorded to them as persons, coming to be treated merely as means and not also as ends in themselves. I think we all know the type who behaves like this: the kind of person who sees everyone else as existing only to help him achieve his goals, never accepting that others might have goals of their own.

Personality traits of this kind are largely a matter of habit. To begin with we become habituated to treating our technologies in this way, and this then readily extends to the treatment of people whom we use as means. Ultimately, if it becomes endemic in the population, we often find that it results in a dystopia, where a whole class of people are treated merely as technologies: the workers in Fritz Lang's film *Metropolis* (1927), or in Chaplin's *Modern Times* (1936). The central idea, then, is that there is a kind of slippery slope here, from the way we treat technologies, to the way we treat people. Largely as a matter of habit, we move readily from treating technologies merely as means to treating people merely as means. And we should cultivate our personality traits to make sure that we do not slide down this slippery slope, and, in order to do this, we should avoid abusing technologies. Thus we would be wrong to think that abusing technologies, in the privacy of one's own home or workplace, is a harmless activity.

There is an analogy here with Kant's discussion of our duties with regard to non-human animals. (In what follows I am much indebted to the discussion in (Korsgaard 2004).) Kant's idea was that we tend to mistake our duties with regard to non-human animals for a duty *towards* those animals – a duty that that we have in virtue of those animals having some kind of call on us. (Kant called this an “amphiboly”.) Kant thought that the only kind of thing that we have duties towards is human beings (ourselves and others) as rational animals. He maintained, in contrast, that we have duties *with regard to* other animals; the mistake (the amphiboly), he thought, was to think that we have duties *towards* them. Non-human animals, Kant thought, are “analogues” of humanity, and our duty is not to them, but to ourselves, “to cultivate our duties to humanity” by acting and feeling dutifully *in respect of* non-human animals. Kant put it thus:

With regard to the animate but non-rational part of creation, violent and cruel treatment of animals is . . . intimately opposed to a human being's duty to himself, and he has a duty to refrain from this; for it dulls his shared feeling of their suffering and so weakens and gradually uproots a natural predisposition that is very serviceable to morality in one's relation with other people. The human being is authorized to kill animals quickly (without pain) and to put them to work that does not strain them beyond their capacities (such work as he himself must submit to). But agonizing physical experiments for the sake of mere speculation, when the end could also be achieved without these, are to be abhorred. – Even gratitude for the long service of an old horse or dog (just as if they were members of the household) belongs *indirectly* to a human being's duty *with regard to* these animals; considered as a *direct* duty, however, it is always only a duty of the human being *to himself* (Kant 1797/1996, p. 443), cited in part in Korsgaard (2004, pp. 90–91).

Now, I do not want to consider whether or not Kant's views about non-human animals is correct, or whether an alternative view (such as that of Peter Singer) is to be preferred, a view that ascribes rights to non-human animals, so that we have consequent duties *towards* them and not merely *with regard to* them. For we can reject



Kant's views about non-human animals, but still insist on the correctness of the parallel view in relation to technologies. So technologies have no rights and we have no consequent duties *towards* them. But we have duties *in respect of* technologies. This is the duty to ourselves to cultivate our personality traits in respect of them, because acting in accordance with this duty cultivates our acting dutifully towards people, whom we should always treat as ends.

Recall here, though, that I am merely arguing that this duty with regard to technologies extends only to not treating them badly or abusing them in the various ways I have been discussing. It does not extend to the kinds of positive duties that are involved in respect for people – nor, indeed, to the gratitude for long service that we accord to the horse or the dog! So the range of personality traits to focus on will include, for example, curtailing irritability and short-temperedness, and not on, for example, developing gratitude and generosity.

It might be complained that what I am proposing is motivationally paradoxical, in the sense that I am advocating that we should be motivated to treat technologies as if they have non-instrumental value in spite of knowing that they do not have such a value, and that we should do so in order to avoid a slide down the slippery slope. The paradox, according to the complaint, is that the motivating reasons for adopting the practice are in fact external to the practice whilst we are supposed to treat them as if they are internal – as if technologies really do have non-instrumental value so that our duties are towards them. But the motivational paradox is not as tight as the complaint suggests. Consider, for example, how one might begin jogging in the morning in order to lose weight, but one appreciates that in order to do this every morning one must enjoy running for its own sake. It sounds paradoxical to say “I should enjoy running for its own sake in order to lose weight”, but the motivational pattern is clear enough. Many of our motivations for practicing certain kinds of behaviour begin as external, but in the knowledge that the best way of keeping up the practice is for the motivations to become internal to the practice.

A further complaint against what I am suggesting is that my claim rests on the idea that there really is a slippery slope here, and this is open to question. Indeed, there are some slippery slope arguments that are problematic, but this is not one such. In his paper “What slopes are slippery?”, Bernard Williams made the distinction between two types of slippery slope argument: the “arbitrary result” argument; and the “horrible result” argument. The latter relies both on the argument that there is “no point at which one can non-arbitrarily get off the slope once one has got on to it”, and on the further argument “that there is a clearly objectionable practice to which the slope leads” (Williams 1995, p. 213). As an example of the first, Williams considers the claim that the extension of some kind of married person's tax relief, from couples who are legally married to some other couples, would put one on a slippery slope where any cut off point in the relief would end up as arbitrary. As an example of the second, Williams mentions the argument against *in vitro* fertilization of human ova.

My argument is of the second kind: the “horrible result” is the failure to treat other people as they ought to be treated: not merely as means but also as ends in

themselves. And we come to do this, so the argument goes, as a consequence of abusing technologies in various ways. So the argument against abusing technologies is, in this sense, consequentialist. From this it will be evident that it is necessary for this argument to go through that there be a plausible *psychological* slippery slope, from the abuse of technologies to the abuse of people, as, for example, there is a psychological slippery slope for the alcoholic in moving from one drink to one drink too many (Williams 1995, p. 218). In respect of my argument, the psychological slippery slope involves interesting issues concerning the relation between personality traits, moods, and emotions (Goldie 2000; Goldie 2004). Consider the person who starts the day before going to work abusing his computer, his mobile phone, and various other technologies. His emotion towards these things is one of anger – anger that they will not work and interact with him as they ought. These emotions put him in an irritable mood – one where he is prone to get angry at other things that will not do as he wants: towards his children for not eating their breakfast; and then later in the morning towards the man in the ticket office for not dealing with his request as he thinks appropriate. And these emotions and moods, over time, consolidate into a personality trait: into the disposition to get angry and to abuse people in general as well as technologies in general. The psychological slippery slope, then, does not involve the risk that there is “some motive . . . to move from one step to the next” (Williams 1995, p. 218). The risk, rather, is that one becomes habituated in feeling and behaving a certain way, and thus one unthinkingly moves, out of habituation grounded ultimately in a personality trait, from one step to the next.

Finally, there might be a concern that my idea, that we should behave with respect towards technologies by not abusing them, runs the risk of putting us on a different slippery slope: this time from treating technologies with respect by not abusing them, to treating them with respect by treating them just as we would treat human beings. It might be said, in support of this concern, that a similar slippery slope can arise in our treatment of animals: the animal lover sometimes comes to treat non-human animals and humanity with *equal* respect, or even, at the extreme, with *more* respect than humans, as perhaps is sometimes found with animal rights campaigners, who abuse, terrorise, or even kill their fellow human beings in order to protect other animals. Treating animals with respect ought not to turn into treating them just as we treat humans. And, of course, the same point applies to technologies – a fortiori one might reasonably think. I think we can accept that there are some grounds for this concern with small children, as evidenced by the tyranny that Tagamochi toys can have over their lives. But this particular slippery slope argument fails, because there is not a genuine *psychological* slippery slope here. Negative, abusive behaviour of the kind I have been concerned with is habitual and characteristically not reason-based, so that one can all too easily slide from abusing technology, to abusing non-human animals, and then to abusing people. In contrast, positive, caring behaviour is characteristically reason-based and not habitual, so there is no reason to think that this slippery slope is a concern for most adults. Just as most of us are able to distinguish between our positive duties with regard to

non-human animals from our duties towards humans, I think we can do the same with technologies. Things are different, though, with our bad habits.

## 7 Conclusion

Freud's remarks in 1930 show us that, in a sense, there is nothing new in our relation to technologies: in spite of the advances, they continue to give us difficulties at times. And yet, in another sense, there *is* something new. Today, we *interact* with many technologies in ways that we did not in Freud's day: we interact with robots, with avatars, with androids, with EOTs, and with ECAs. Like non-human animals, but in a different way, they have become, to use Kant's term, *analogues* of humanity. And, because of this, there is now a particularly slippery psychological slope from the abusive ways in which we can treat these technologies to the abusive ways in which we come to treat humanity. This slope is to be avoided, and the way to do so is to cultivate our personality traits so that we treat technologies with respect, to the extent of not abusing or otherwise behaving badly towards them.

Finally, I should say something about the relationship between the normative and the empirical issues in this chapter. I said that I would focus mainly on normative questions rather than empirical ones, but in the end it is important to accept that my slippery slope argument depends on certain facts about human psychology, and it is, in just that sense, empirical.

**Acknowledgments** My thanks to Sabine Roeser for her support as editor of this volume, and to Roddy Cowie and others for their contribution to, and discussion of, work on which this chapter is based, and to HUMAINE (Human-Machine Interaction Network on Emotion), a Network of Excellence in the EU's Sixth Framework Programme (Contract no. 507422).

## References

- Aristotle. *Nicomachean ethics*, Translated by W. D. Ross and revised by J. O. Urmson. In *The Complete Works of Aristotle: The Revised Oxford Translation, Vol. 2*. J. Barnes, ed., Princeton: Princeton University Press.
- Bartneck, C. et al. 2006. To kill a robot. *Proceedings of the Workshop on Misuse and Abuse of Interactive Technologies in cooperation with the Conference on Human Factors in Computing Systems CHI2006* Montreal, Canada.
- Bartneck, C., and J., Hu. 2008. Exploring the abuse of robots. *Interaction Studies – Social Behaviour and Communication in Biological and Artificial Systems* 9: 415–433.
- Block, N. 1997. Biology versus computation in the study of consciousness. *Behavioral and Brain Sciences* 20: 159–165.
- de Angeli, A., S., Brahmam, P., Wallis, and A., Dix (2006). Misuse and abuse of interactive technologies. *Proceedings of the Workshop on Misuse and Abuse of Interactive Technologies in cooperation with the Conference on Human Factors in Computing Systems CHI2006* Montreal, Canada.
- Dennett, D. 1996. *Kinds of Minds: Toward an Understanding of Consciousness*. New York: Basic Books.
- Freud, S. 1930/1985. Civilization and its discontents. In *The Penguin Freud Library, Vol. 12: Civilization, Society and Religion*. S. Freud, ed., 243–340, London: Penguin.

- Goldie, P. 2009. Thick concepts and emotion. In *Reading Bernard Williams*. D. Callcut, ed., 94–109, London: Routledge.
- Goldie, P. 2004. *On Personality*. London: Routledge.
- Goldie, P. 2000. *The Emotions: A Philosophical Exploration*. Oxford: Clarendon Press.
- Hatzimoyisis, A. 2003. Sentimental value. *Philosophical Quarterly* 53: 373–379.
- Hursthouse, R. 1991. Arational actions. *The Journal of Philosophy* 88: 57–68.
- Kant, I. 1785/1964. *Groundwork of the Metaphysics of Morals*. Translated by H. J. Paton, New York: Harper & Row.
- Kant, I. 1790/1953. *The Critique of Judgement*. Translated by J. C. Meredith, Oxford: Oxford University Press.
- Kant, I. 1797/1996. *The Metaphysics of Morals*. Translated by M. Gregor, Cambridge: Cambridge University Press.
- Korsgaard, C. 2004. Fellow creatures: Kantian ethics and our duties to animals. *Tanner Lectures on Human Values* 24: 79–110.
- MacDorman, K., and H., Ishiguro. 2006. The uncanny advantage of using androids in cognitive and social science research. *Interaction Studies* 7: 297–337.
- Milgram, S. 1974. *Obedience to Authority: An Experimental View*. New York: Harper and Row.
- Mori, M. 1970. The uncanny valley. *Energy* 7: 33–35.
- Nagel, T. 1974. What is it like to be a bat? *The Philosophical Review* 83: 435–450.
- Nass, C., and B., Reeves. 1996. *The Media Equation: How People Treat Computers*. Cambridge: Cambridge University Press.
- Picard, R. 2002. What does it mean for a computer to ‘have’ emotions? In *Emotions in Humans and Artefacts*. R. Trapp, P. Pettaand, and S. Payr, eds., Cambridge, MA: MIT Press.
- Radford, C. 2001. Paradoxes of emotion and fiction. *The Philosophical Review* 110: 617–620.
- Rosalia, C., R., Menges, I., Deckers, and C., Bartneck (2005). Cruelty towards robots. *Robot Workshop – Designing Robot Applications for Everyday Use*, Göteborg.
- Scarre, G. 1981. Kant on free and dependent beauty. *British Journal of Aesthetics* 21: 351–362.
- Singer, P. 1977. *Animal Liberation: Towards an End to Man’s Inhumanity to Animals*. Boulder, CO: Paladin.
- Slater, M., A., Antley, A., Davison, D., Swapp, C., Guger et al. 2006. A virtual reprise of the Stanley Milgram obedience experiments. *PLoS ONE* 1: 1.
- Stocker, M. 1976. The schizophrenia of modern ethical theories. *The Journal of Philosophy* 73: 453–466.
- Williams, B. 1995. Which slopes are slippery? In *Making Sense of Humanity and Other Philosophical Papers*. B. Williams, ed., 213–223, Cambridge: Cambridge University Press.